Speech Processing 2010/2011

2nd Test

April 18th 2011

Please identify this form with your name and student number in the reserved space at the bottom. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

1. Classify as True (T) or False (F)
   (a) MELP coders differ from typical vocoders in that the source is not either voiced or unvoiced only.
   (b) In subband coders, QMF filters are successively applied to the upper subband in order to achieve a fair approximation to the critical band concept.
   (c) The minimum delay in speech coders is 2 frames.
   (d) The adaptive codebook in CELP coders plays a similar role to the long-term predictor.
   (e) Vector quantization of short-term prediction coefficients is less efficient than scalar quantization for a given bit rate, in terms of the achieved distortion.
2. (a) What was the bit rate range that used to be known as the coding gap?
   (b) What is the typical number of poles in a low bit rate coder using the linear prediction model?
3. Give examples (one per class) of
   (a) a coder which uses nearest neighbor frequency tracking
   (b) an early parametric coder
   (c) a coder which transmits the amplitudes of grids of pulses
   (d) a coder which uses a quantization step proportional to the signal value
   (e) a coder which uses structured codebooks for the excitation
4. Take as a starting point a CELP coder operating at 8 kbps, based on frames of 5ms. Show the bit assignment for each type of parameter at this bit rate and how you would increase it to 8.2 kpbs.
5. Classify as True (T) or False (F)
   (a) Formant synthesizers such as Klatt's use a cascade of a ressonator and antiressonator to simulate nasal sounds.
   (b) ToBI models are used for duration modeling.
   (c) In a neural network trained for GtoP, typically 50% of the aligned letter-phone data would be reserved for testing.
   (d) CART-based approaches for prosodic phrasing may be trained on annotated texts, without the corresponding wav files.
   (e) TD-PSOLA synthesis is computationally more expensive than MBROLA.
   (f) Data-driven approaches are more appropriate than rule-based approaches for converting digit sequences (as in natural numbers) into letter sequences.
6. The concatenation cost is a weighted sum of the differences between several features of two potential speech units. Indicate these features.

7. Which is the consonant needing a greater number of rules in Portuguese?

8. Name one of the parameters of articulatory synthesizers that allows the production of nasal vowels.

9. The word to be synthesized is *cavalo*.

   (a) The available corpus has the words *crocodilo*, *gorila*, *macaco*, *galinha*, *avestruz*, *vaca*, *pombo*. Indicate which phone sequences you will select from which words to concatenate the target word, using the fewest sequences. If you miss any phone sequence indicate which one. Use # as a symbol for silence.

   (b) The prosody generation module computed a target duration of 100ms for the fricative sound. The left sub-unit has a duration value of 40ms and the right sub-unit has a duration value of 50ms. Indicate the target durations for these two sub-units of the fricative.

10. Write two pronunciations for the following foreign names and acronyms. The first pronunciation should be the one typically heard in Portuguese media. The second pronunciation is the one obtained using the grapheme-to-phone rules for the common lexicon in European Portuguese. Both should use only the phonetic SAMPA symbols of this language.

    (a) Manchester United

    (b) PME

    (c) REFER

11. Using the following syntax
    a → b / c .. d
    describe the simplified rules for grapheme *s*, where c and d may be graphemes (e.g.: a, b, etc.), classes of graphemes (vowels, consonants, etc.), the word boundary (#), or any grapheme (*). You may use the symbol *0* to mark phonemic nulls, and the symbols | and ( ) to mark disjunction between several graphemes (e.g.: a | b | c). The rules are applied in order, until one matches the context and, in this case, the following rules are not applied. The "simpliflied" rules do not need to contemplate compound words, words of foreign origin, or even words with double *s*. In fact, they should only account for the cases depicted in the examples below (in alphabetical order):
    alsa, arsénio, caros, caso, castelo, manso, osga, seta

Test 2 - Answers

Name:	
Number:	

1. (1.5 val.) Indicate T or F:

a	b	c	d	e

2. (1.2 val.)

a	
b	

3. (2.5 val.)

a	
b	
c	
d	
e	

4. (2.3 val.)

	8 kpbs	8.2 kpbs
long-term predictor		
short-term predictor		
excitation		

5. (1.8 val.) Indicate T or F:

a	b	c	d	e	f

6 to 9 (1.2/1.0/1.0/2.5 val.)

6	
7	
8	
9	

10. (2.0 val.)

transc	Manchester United	PME	REFER
transc1			
transc2			

The answer to the last question (11) should be written in the next page (3 val.).